

Desafio de Análise de Dados Agrícolas com SQL Avançado

Objetivo do Projeto

O objetivo deste desafio é avaliar a capacidade de utilização da ferramenta de consulta de dados, SQL. As perguntas propostas devem ser respondidas de maneira eficaz e otimizada.

Fontes de Dados

Os dados utilizados vêm da [Pesquisa Agrícola Municipal](#) e incluem as seguintes colunas:

- ano: Ano da coleta de dados.
- sigla_uf: Unidade da Federação.
- id_municipio: Código IBGE do município.
- produto: Tipo de lavoura.
- area_destinada_colheita: Área total destinada à colheita (ha).
- area_colhida: Área efetivamente colhida (ha).
- quantidade_produzida: Quantidade total produzida (toneladas).
- rendimento_medio_producao: Produção média por hectare (kg/ha).
- valor_producao: Valor total da produção em mil reais.

Ferramentas Utilizadas

SQL Avançado: linguagem para fazer as consultas e cálculos necessários.

- CTEs e Joins: Para integrar diferentes tabelas e criar agregações.
- Window Functions: Para cálculos acumulativos e médias móveis.
- Subconsultas: Para cálculos dinâmicos.

Big Query: serviço de análise de dados do Google Cloud Platform (GCP) que permite armazenar e analisar grandes quantidades de dados.

Importando a base para o Big Query

SELECT

```
dados.ano AS ano,
dados.sigla_uf AS sigla_uf,
diretorio_sigla_uf.nome AS sigla_uf_nome,
dados.id_municipio AS id_municipio,
diretorio_id_municipio.nome AS id_municipio_nome,
dados.produto AS produto,
dados.area_destinada_colheita AS area_destinada_colheita,
dados.area_colhida AS area_colhida,
dados.quantidade_produzida AS quantidade_produzida,
dados.rendimento_medio_producao AS rendimento_medio_producao,
dados.valor_producao AS valor_producao
```

FROM `basedosdados.br_ibge_pam.lavoura_permanente` AS dados

```
LEFT JOIN (SELECT DISTINCT sigla,nome FROM
`basedosdados.br_bd_diretorios_brasil.uf`) AS diretorio_sigla_uf
ON dados.sigla_uf = diretorio_sigla_uf.sigla
LEFT JOIN (SELECT DISTINCT id_municipio,nome FROM
`basedosdados.br_bd_diretorios_brasil.municipio`) AS diretorio_id_municipio
ON dados.id_municipio = diretorio_id_municipio.id_municipio
```

Análise Exploratória de Dados (EDA)

1. **Produção do setor agrícola**
 - a. Como a produção total de lavouras permanentes se distribuiu entre os estados e municípios?
 - b. Quais produtos dominam a produção agrícola em cada região?
 - c. A produção aumentou ou diminuiu ao longo dos anos?
2. **Produtividade do setor agrícola**
 - a. Existe alguma relação entre o aumento da área colhida e a produtividade do produto?
 - b. Qual a produtividade média varia entre os municípios?
3. **Desempenho do setor agrícola**
 - a. Como o desempenho das lavouras varia de acordo com a região e produto ao longo dos anos?
 - b. Como a eficiência agrícola (produtividade média por hectare) evolui ao longo dos anos e como ela se distribui entre os estados e os produtos?
4. **Projeções do setor agrícola**
 - a. Quais são as projeções para a produção agrícola nos próximos anos?

Análise Exploratória de Dados (EDA)

Produção do setor agrícola

Como a produção total de lavouras permanentes se distribuiu entre os estados e municípios?

```
WITH producao AS (  
  SELECT  
    sigla_uf,  
    id_municipio,  
    ROUND(SUM(quantidade_produzida)) AS prod_municipio,  
    SUM(SUM(quantidade_produzida)) OVER (PARTITION BY sigla_uf) AS prod_estado,  
    SUM(SUM(quantidade_produzida)) OVER () AS prod_brasil  
  FROM principal.pam  
  GROUP BY sigla_uf, id_municipio  
  LIMIT 10  
)  
  
SELECT  
  sigla_uf,  
  id_municipio,  
  ROUND(100 * prod_municipio / prod_estado, 2) AS perc_municipio,  
  ROUND(100 * prod_estado / prod_brasil, 2) AS perc_estado,  
  RANK() OVER (PARTITION BY sigla_uf ORDER BY prod_municipio DESC) AS  
  ranking_estado  
FROM producao  
ORDER BY sigla_uf, ranking_estado, prod_municipio ASC;
```

Resultados da consulta [Salvar resultados](#)

Informações do job	Resultados	Gráfico	JSON	Detalhes da execução	Gráfico de execução
Linha	sigla_uf	id_municipio	perc_municipio	perc_estado	ranking_estado
1	CE	2302602	2.18	3.29	1
2	CE	2301901	0.41	3.29	2
3	CE	2304905	0.04	3.29	3
4	CE	2300804	0.03	3.29	4
5	CE	2311355	0.01	3.29	5
6	CE	2301851	0.01	3.29	6
7	MA	2111573	0.08	0.45	1
8	MA	2104651	0.04	0.45	2
9	RO	1100189	1.73	0.43	1
10	RS	4320230	0.02	3.77	1

Resultados por página: 50 1 – 10 de 10

Quais produtos dominam a produção agrícola em cada região?

```
SELECT
  sigla_uf,
  produto,
  ROUND((SUM(quantidade_produzida) * 100.0) /
    SUM(SUM(quantidade_produzida)) OVER (PARTITION BY sigla_uf), 2)
    AS percentual_producao_uf
FROM principal.pam
GROUP BY sigla_uf, produto
ORDER BY percentual_producao DESC;
```

Resultados da consulta					
Informações do job		Resultados	Gráfico	JSON	Detalhes da execução
Linha	sigla_uf	produto	percentual_producao_uf		
1	SP	Laranja	92.65		
2	SE	Laranja	88.57		
3	TO	Banana (cacho)	81.86		
4	AC	Banana (cacho)	76.72		
5	RJ	Laranja	72.98		
6	RR	Banana (cacho)	72.15		
7	PI	Caju	64.68		
8	CE	Caju	59.11		
9	AM	Banana (cacho)	58.2		
10	MT	Banana (cacho)	55.04		

A produção aumentou ou diminuiu ao longo dos anos?

```
WITH consulta AS (
  SELECT
    ano,
    ROUND(SUM(quantidade_produzida)) AS producao
  FROM principal.pam
  GROUP BY ano
  ORDER BY ano DESC
)
```

```
SELECT
  ano,
  producao,
```

```
LAG(producao) OVER (ORDER BY ano) AS producao_anterior,
ROUND((producao - LAG(producao) OVER (ORDER BY ano)) / LAG(producao) OVER (ORDER
BY ano) * 100, 2) AS crescimento_percentual
FROM consulta
ORDER BY ano DESC;
```

Resultados da consulta					
Informações do job		Resultados	Gráfico	JSON	Detalhes da execução
Linha	ano	producao	producao_anterior	crescimento_percentual	
1	2022	48304825.0	47072280.0	2.62	
2	2021	47072280.0	48076282.0	-2.09	
3	2020	48076282.0	46722270.0	2.9	
4	2019	46722270.0	46842797.0	-0.26	
5	2018	46842797.0	44684821.0	4.83	
6	2017	44684821.0	44176571.0	1.15	
7	2016	44176571.0	44331373.0	-0.35	
8	2015	44331373.0	43906120.0	0.97	
9	2014	43906120.0	44370588.0	-1.05	
10	2013	44370588.0	45121841.0	-1.66	

Produtividade do setor agrícola

Existe alguma relação entre o aumento da área colhida e a produtividade do produto?

```
WITH dados_produtividade AS (
  SELECT
    produto,
    ano,
    SUM(area_colhida) AS area_total_colhida,
    SUM(quantidade_produzida) / NULLIF(SUM(area_colhida), 0) AS produtividade_media
  FROM principal.pam
  GROUP BY produto, ano
  HAVING COUNT(*) > 1 -- Garante que há mais de um ponto de dados para calcular a
  correlação
)

SELECT
  produto,
  ROUND(CORR(area_total_colhida, produtividade_media), 2) AS
  correlacao_area_produtividade
```

```
FROM dados_produtividade
GROUP BY produto
HAVING correlacao_area_produtividade IS NOT NULL
ORDER BY correlacao_area_produtividade DESC;
```

Resultados da consulta

Salvar resultados

Informações do job

Resultados

Gráfico

JSON

Detalhes da execução

Gráfico de execução

Linha	produto	correlacao_area_pro
1	Borracha (látex coagulado)	0.98
2	Maçã	0.92
3	Dendê (cacho de coco)	0.87
4	Limão	0.84
5	Uva	0.8
6	Mamão	0.75
7	Tangerina	0.67
8	Coco-da-baía*	0.53
9	Manga	0.35
10	Borracha (látex líquido)	0.33

Resultados por página: 501 – 38 de 38

Qual a produtividade média varia entre os municípios?

Essa query permite categorizar municípios em pequeno, médio ou grande produtor com base na proporção de cada em relação à área colhida. A divisão da área colhida foi feita em três partes iguais:

Pequeno: Área colhida menor que 1/3 do intervalo total acima do mínimo.

Médio: Área colhida entre 1/3 e 2/3 do intervalo total.

Grande: Área colhida acima de 2/3 do intervalo total.

```
WITH area_produtiva AS (
  SELECT
    sigla_uf,
    id_municipio,
    produto,
    ano,
    ROUND(SUM(area_colhida), 2) AS area_colhida,
    ROUND(CASE WHEN SUM(area_colhida) = 0 THEN 0
                ELSE SUM(quantidade_produzida) / SUM(area_colhida) END, 2) AS
    produtividade
  FROM principal.pam
  GROUP BY sigla_uf, id_municipio, produto, ano
),
```

```
classificacao_produtores AS (  
  SELECT  
    sigla_uf,  
    id_municipio,  
    produto,  
    produtividade,  
    area_colhida,  
    CASE  
      WHEN area_colhida < (SELECT MIN(area_colhida) FROM area_produtiva) + (SELECT  
(MAX(area_colhida) - MIN(area_colhida)) / 3 FROM area_produtiva) THEN 'Pequeno'  
      WHEN area_colhida >= (SELECT MIN(area_colhida) FROM area_produtiva) + (SELECT  
(MAX(area_colhida) - MIN(area_colhida)) / 3 FROM area_produtiva)  
        AND area_colhida < (SELECT MIN(area_colhida) FROM area_produtiva) + 2 *  
(SELECT (MAX(area_colhida) - MIN(area_colhida)) / 3 FROM area_produtiva) THEN  
'Médio'  
      ELSE 'Grande'  
    END AS categoria_produtores  
  FROM area_produtiva  
)  
  
SELECT  
  categoria_produtores,  
  ROUND(AVG(produtividade), 2) AS produtividade_media  
FROM classificacao_produtores  
GROUP BY categoria_produtores  
ORDER BY categoria_produtores;
```

Resultados da consulta

Informações do job		Resultados	Gráfico	JSON	Det
Linha	categoria_produtores	produtividade_media			
1	Grande	0.5			
2	Médio	14.86			
3	Pequeno	15.3			

Desempenho do setor agrícola

Como o desempenho das lavouras varia de acordo com a região e produto ao longo dos anos?

```
WITH desempenho_regional AS (  
  SELECT  
    sigla_uf,  
    id_municipio,  
    produto,  
    ROUND(SUM(area_colhida)) AS area_colhida,  
    ROUND(SUM(quantidade_produzida)) AS producao,  
    CASE  
      WHEN SUM(area_colhida) > 0 THEN ROUND(SUM(quantidade_produzida) /  
SUM(area_colhida), 2)  
      ELSE 0  
    END AS rendimento_medio  
  FROM principal.pam  
  GROUP BY sigla_uf, id_municipio, produto  
)  
SELECT  
  sigla_uf,  
  id_municipio,  
  produto,  
  area_colhida,  
  producao,  
  rendimento_medio  
FROM desempenho_regional  
ORDER BY sigla_uf, rendimento_medio DESC;
```

Resultados da consulta

Salvar resultados

Abrir em

Informações do job	Resultados	Gráfico	JSON	Detalhes da execução	Gráfico de execução	
Linha	sigla_uf	id_municipio	produto	area_colhida	producao	rendimento_medio
1	AC	1200708	Laranja	431.0	28626.0	66.42
2	AC	1200104	Laranja	403.0	25050.0	62.16
3	AC	1200401	Laranja	2036.0	87795.0	43.12
4	AC	1200054	Laranja	99.0	3895.0	39.34
5	AC	1200252	Laranja	279.0	10660.0	38.21
6	AC	1200450	Laranja	1330.0	48304.0	36.32
7	AC	1200450	Manga	193.0	6365.0	32.98
8	AC	1200302	Laranja	641.0	19358.0	30.2
9	AC	1200500	Laranja	1187.0	35167.0	29.63
10	AC	1200179	Laranja	293.0	8391.0	28.64

Resultados por página: 501 – 50 de 210330<<>>

Como a eficiência agrícola (produtividade média por hectare) evolui ao longo dos anos e como ela se distribui entre os estados e os produtos?

```
WITH eficiencia AS (  
    SELECT  
        ano,  
        sigla_uf,  
        produto,  
        ROUND(SUM(quantidade_produzida) / NULLIF(SUM(area_colhida), 0), 2) AS  
        eficiencia_media  
    FROM principal.pam  
    GROUP BY ano, sigla_uf, produto  
)  
  
SELECT  
    ano,  
    sigla_uf,  
    produto,  
    eficiencia_media,  
    LAG(eficiencia_media) OVER (PARTITION BY sigla_uf, produto ORDER BY ano) AS  
    eficiencia_anterior,  
    CASE  
        WHEN LAG(eficiencia_media) OVER (PARTITION BY sigla_uf, produto ORDER BY ano)  
        IS NOT NULL  
        THEN ROUND(((eficiencia_media - LAG(eficiencia_media) OVER (PARTITION BY  
sigla_uf, produto ORDER BY ano)) /  
        LAG(eficiencia_media) OVER (PARTITION BY sigla_uf, produto ORDER BY ano)) *  
100, 2)  
        ELSE NULL  
    END AS var_percentual_eficiencia  
FROM eficiencia  
WHERE eficiencia_media IS NOT NULL  
ORDER BY sigla_uf, produto, ano;
```

Resultados da consulta

Salvar resultados

Abrir em

Informações do job	Resultados	Gráfico	JSON	Detalhes da execução	Gráfico de execução	
Linha	ano	sigla_uf	produto	eficiencia_media	eficiencia_anterior	var_percentual_efic
1	1974	AC	Abacate	16.82	null	null
2	1975	AC	Abacate	18.23	16.82	8.38
3	1976	AC	Abacate	17.15	18.23	-5.92
4	1977	AC	Abacate	18.1	17.15	5.54
5	1978	AC	Abacate	18.7	18.1	3.31
6	1979	AC	Abacate	18.66	18.7	-0.21
7	1980	AC	Abacate	20.57	18.66	10.24
8	1981	AC	Abacate	18.62	20.57	-9.48
9	1982	AC	Abacate	20.71	18.62	11.22
10	1983	AC	Abacate	12.73	20.71	-38.53

Projeções do setor agrícola

Quais são as projeções para a produção agrícola nos próximos anos?

```
WITH consulta AS (  
  SELECT  
    ano,  
    ROUND(SUM(quantidade_produzida)) AS producao  
  FROM principal.pam  
  GROUP BY ano  
) ,  
  
crescimento_anual AS (  
  SELECT  
    ano,  
    producao,  
    (producao - LAG(producao) OVER (ORDER BY ano)) / LAG(producao) OVER (ORDER BY  
ano) AS taxa_crescimento  
  FROM consulta  
) ,  
  
media_crescimento AS (  
  SELECT AVG(taxa_crescimento) AS taxa_crescimento  
  FROM crescimento_anual  
  WHERE taxa_crescimento IS NOT NULL  
)  
  
SELECT  
  (SELECT MAX(ano) FROM consulta) + n AS ano_projetado,  
  ROUND((SELECT MAX(producao) FROM consulta) * POWER(1 + (SELECT taxa_crescimento  
FROM media_crescimento), n), 0) AS producao_projetada  
FROM UNNEST(GENERATE_ARRAY(1, 5)) AS n  
ORDER BY ano_projetado;
```

Resultados da consulta

Informações do job	Resultados	Gráfico	JSON	Detalhes da execução	Gráfico de execu
Linha	ano_projetado	producao_projetada			
1	2023	134542948.0			
2	2024	136479398.0			
3	2025	138443719.0			
4	2026	140436311.0			
5	2027	142457583.0			